Review of Issues In Intelligence Test Use In Alberta Schools

Special Educational Services

Alberta
EDUCATION

REVIEW OF ISSUES IN INTELLIGENCE TEST USE IN ALBERTA SCHOOLS

Special Educational Services

Alberta Education

November 1986

CONTENTS

	Page
INTRODUCTION	1
Purpose of Review	1
Rationale	1
Methodology	1
A BRIEF REVIEW OF SELECTED LITERATURE	3
The Nature of Intelligence	3
The Measurement of Intelligence	5
Instrument Variables / 6	
Examiner Variables / 12	
Interpretation Variables / 13	
The Rights of the Student	14
Implications of Intelligence Test Results in Decision Making	15
Labelling / 15	
Placement / 16	
Special Populations	16
Minority Students / 16	
English As A Second Language (ESL) / 17	
Native Children / 19	
Learning Disabled / 20	
Behaviorally Disordered / 20	
Physically Disabled / 21	
Gifted / 21	
Summary	23
SURVEY OF ALBERTA SCHOOL JURISDICTIONS	24
STATUS OF POLICY ON INTELLIGENCE TESTS IN CANADA	25
POLICY ON INTELLIGENCE TESTS IN THE UNITED STATES	27
SUMMARY OF FINDINGS	28
GLOSSARY	29
REFERENCES	31

INTRODUCTION

PURPOSE OF REVIEW

In March, 1984 the Special Educational Services Branch of Alberta Education undertook a review to identify and verify issues involved in the use of intelligence tests in Alberta schools. Specifically, the purpose of the review included:

1. The identification of issues involved in the use of intelligence tests in schools, particularly those issues that pertain to exceptional individuals, including the gifted.

2. The determination of the ways in which intelligence tests are used in Alberta schools.

3. The development and provision of a literature review regarding identified issues surrounding intelligence testing.

4. The development and provision of a summary report.

RATIONALE FOR REVIEW

Intelligence tests are utilized regularly in Alberta schools as one guidepost for program and placement decisions. These decisions impact on the future educational, career and personal/social development of children, particularly those who are referred to as exceptional. Intelligence tests are generally assumed to measure "intelligence" and intelligence is assumed to be of a fixed quality and quantity. For some time these assumptions have been challenged. Nevertheless, educators continue to place a high degree of trust in the score resulting from the administration of a single intelligence test. The present review arises from presentations to Alberta Education officials by school jurisdictions and various interest groups regarding the appropriate use of intelligence tests.

METHODOLOGY

Information was obtained from various sources including: a critical review of the related literature; presentations by and discussions with six invited educators; a survey of Alberta school jurisdictions to determine current practice; and, a brief survey of other Canadian provinces and selected states in the United States to determine common practice and policy regarding the use of intelligence tests.

An advisory committee was formed to assist in the review and to provide monitoring and advice to the Alberta Education staff which comprised the working team. The advisory committee was made up of the following school jurisdiction and university representatives:

- Mr. Don Dawson, Consultant/School Psychologist, Lethbridge Public School District
- Dr. Bruce Handley, Pupil Personnel Coordinator, Red Deer Public School District
- Dr. Sharon Robertson, Associate Professor, Department of Educational Psychology, University of Calgary
- Mrs. Marian Stelmaschuk, Teacher/Consultant, Edmonton Public School District

Representing Alberta Education on the advisory committee were:

- Mr. Fred French, Coordinator of Guidance and Career Development, Special Educational Services Branch (Chairperson)
- Dr. David Wasserman, Assessment Resources Officer, Student Evaluation Branch

Executive secretary and initial research duties were conducted by Ms. Donna Patterson, an intern with Alberta Education at the time of the review.

Alberta Education expresses its appreciation to the advisory committee, presenters and responders who made the review possible. Recognition and appreciation is also extended to the following professional educators for their contributions in specific topic areas:

1. Dr. Agnes Yu, Edmonton Public School District
   - Special Problems of Children for whom English is a Second Language (ESL), and IQ Testing
2. Dr. Philip Vernon, University of Calgary
   - The Nature of Intelligence
3. Dr. Roy Brown, University of Calgary
   - The Use of IQ Tests with Severely Handicapped Children and Adults.
4. Dr. Lillian White, University of Alberta
   - The Use of IQ Tests for Diagnostic Program Purposes

5. Mr. Bill Rawlusyk, Director Pupil Personnel Services, Calgary R.C.S. School District
   - The Use of IQ Tests with Low Functioning Children and from an Administrative Viewpoint.

6. Dr. Marilyn Samuels, Executive Director, Calgary Learning Centre
   - Alternatives to Intelligence Tests

In addressing the purposes of the study the committee focused on the following questions:

1. What is the current situation in Alberta schools relative to the use of intelligence test scores?

2. What are the perceived issues involved in the use of intelligence tests in schools, by teachers, parents, administrators and other professionals?

3. Do identified issues have merit? That is, are the perceived issues based on fact or opinions resulting from a lack of knowledge concerning the use of intelligence tests?

4. What are the implications of the issues for Alberta Education regarding policy, guidelines and procedures development?

5. What are the implications for school jurisdictions regarding policy, guideline and procedures development?

6. For what, if any, purpose(s) should intelligence tests legitimately be used?

7. How might assessment of intelligence be conducted practically?

A BRIEF REVIEW OF SELECTED LITERATURE

THE NATURE OF INTELLIGENCE

The concept of intelligence as obtained through the repeated use of intelligence tests that hold a high degree of statistical validity and reliability, has become deified. However, as stated by Wesman (1968), intelligence is a hypothetical construct. While aspects of this construct are capable of measurement, intelligence may not exist as a substance or as an entity.

An examination of the historical origins of intelligence testing indicates that the construct "intelligence" has been defined in a variety of ways and that numerous theories have been advanced to explain both intelligence and its development. Examples include the works of Hebb (1948), Ferguson (1954), Bruner (1966), Piaget (1952), Thurstone (1938), Feuerstein (1979) and Kaufman (1983). At present, no substantial consensus exists regarding a single operational definition.

The lack of consensus on the nature of intelligence has implications for the users of intelligence test data. For example, Vernon (1961) and Wesman (1968) have concluded that intelligence tests predict the capacity for learning. However if, as Feuerstein (1979) claims, intelligence tests simply give the examiner a cross-section of what has been learned, then there is the possibility of reaching a false conclusion regarding the potential of the individual to benefit from instruction.

Recently, Sternberg (1984) concluded that educators, in their search for a panacea to provide high-quality education, should not look to tests, whether old or new, to be that panacea. He went on to state that presently known tests can provide only an approximation of a person's intelligence. Further, intelligence, even broadly defined, is only one of the many ingredients in success. Extending this position, Gardner (1983) stated that human cognition or intelligence is concerned with a range of kinds of information or contents that cannot be collapsed into one intellectual heap.

While it is not the intent of this review to resolve the nature of intelligence, the view offered by Staats (1970) has some merit for educators in considering the nature of intelligence and the usefulness of intelligence test data. Staats stated that the nature of intelligence may include the notion of an inherited biological structure that fits individuals for learning. However, the nature of intellectual growth will depend on the nature of, and the opportunities to engage in, learning experiences. The intent in making this statement is not to review the history of genetic endowment versus the effects of the environment on intelligence. Rather, it is to point out that the construct known as intelligence is assumed to have a physiological property that is subject to environmental factors. Indeed, while there is little agreement on the nature of intelligence, there is agreement on three general points with regard to what constitutes intelligence.

First, the nature of intelligence is culturally relative. For example, an intelligent person in one culture may not be perceived as intelligent in another culture (Goodnow, in Warren, 1979). Further, Gage and Berliner (1979) found that western culture viewed abstract thinking as more intelligent than concrete thinking.

Second, De Lisi (1980) stated that factors such as fatigue, motivation and task familiarity are important determinants in how a person performs a task. Ability alone rarely determines performance.

A third area of emerging agreement concerns the challenges to the concepts of intelligence as a fixed immovable state. While there is still a great deal of controversy in this area, a number of recent authors, researchers and theorists such as Sternberg (1984), Feuerstein (1979) and Rose and Rose (1980) argue that intelligence should be viewed as a more dynamic, fluid and multi-trait construct.

THE MEASUREMENT OF INTELLIGENCE

Regardless of how intelligence is conceived and despite some disenchantment with the traditional construct of intelligence, the understanding of the potential of an individual to benefit from instruction is crucial. Even individuals such as Feuerstein (1979) who challenge the traditional static measures of intelligence argue that understanding cognitive functioning is critical to instructional planning. Indeed, Wallace and Larsen (1978) state that one of the major functions of intelligence testing is to acquire information relative to educational program planning. Another function, which is more administrative than clinical, is that of classifying individuals to determine eligibility for placement.

While students typically are referred for intellectual assessment because of problems, one of the most encouraging findings resulting from the administration of an intelligence test is the discovery of skills and strengths that had not been noted previously. In fact, intelligence test results are held to surpass observations, hunches or impressions (Zimmerman and Woo Sam, 1973).

Nevertheless, criticisms of intelligence tests are widespread, as demonstrated by litigations in several United States courts and by the ban on the use of intelligence tests in California (Reschly, 1979). These criticisms relate in part to the confusion of intelligence with the score a student receives on an IQ test, resulting in what Ryan (1973) termed "the illusion of objectivity."

Critics of intelligence tests noted the incidence of minority children who are classified as mildly retarded. This criticism has resulted in the attempt to develop "culture-free" or "culture-fair" intelligence tests (Eells, 1954). Culture-fair tests are supposed to measure performance in terms that are unrelated to any particular socio-economic environment. Accordingly, an environment-free intelligence test could be taken to measure "true" intelligence. In contrast to conventional intelligence tests, culture-fair tests de-emphasize those factors such as speed, item content, and stress on verbal content believed to influence the poorer

performances of minority group children. Culture-fair tests are presented primarily as nonverbal tasks without strict time limits and require neither written nor spoken language on the part of the test taker. Items are selected on the basis of the extent to which they sample knowledge, skills, and experience which are equally common or uncommon to all groups. Various writers such as De Avila & Havassy (1974), Mercer (1973), and Samuda (1975) agree that "culture-free" or "culture-fair" tests are very difficult to construct. Indeed, the construct of "culture-fair" is problematic because it is extremely doubtful if any mental processes exist independent of experience.

With the continuing debate regarding criticisms of intelligence tests, courts have been involved to resolve particular issues. However, law suits in California and Illinois have utilized the same testimony to arrive at diametrically opposed conclusions, disavowing or affirming the use of intelligence tests (Zimmerman and Woo Sam, 1973). Therefore, it is incumbent upon those concerned with the education of students to seek out and utilize not only the best instruments, but also the best techniques in administering, scoring and interpreting measures of intelligence within the context of the strengths and limitations of each test.

The measurement of intelligence or, as it has been more recently conceptualized, cognitive functioning, should be undertaken only for valid reasons. Peterson (1968) held that the only valid reason for spending time in assessment is to generate propositions that are useful in forming decisions relative to benefitting the individual.

Given a valid reason for assessment that relates to the educational program of the student, a brief analysis of the variables that impact on the process of assessment follows.

Instrument Variables

The skillful examiner working under ideal conditions requires precise instruments to obtain the maximum information relative to understanding the mental processing potential of a student.

Generally, intelligence tests fall into two categories. One category is known as the individual intellectual assessment with the test designed to determine the current intellectual functioning of a single individual. A second category is known as the group intellectual assessment with the test designed to measure the current intellectual status of individuals. The adequacy of both types of tests is determined by examining the standardization, validity and reliability of specific tests within these groups. All three characteristics are required.

Standardization is the process of establishing a reference group with which to compare the performance of an individual taking a specific test.

The standardization sample should be large enough to contain a meaningful representative sample of the children to be assessed, including adequate proportions of children of varied ages, racial, geographic, ethnic and socio-economic background.

Reliability refers to the consistency of scores generated under repeated and variable examining conditions. Haphazard fluctuations in an individual's scores over time on the same instrument render the test results meaningless. Consistency over time is expressed as a reliability coefficient with a coefficient of .80 or higher being generally acceptable.

Validity, or the accuracy with which a test measures what it purports to measure, is one of the most critical determinants of a good test. Three types of validity have been identified including content, construct and criterion validity. Criterion validity examines the relationship between the test in question and some type of criterion or outcome. Construct validity examines whether or not the items on a particular test relate to the theoretical constructs that the test purports to measure. Content validity examines whether or not the test items are relative and representative of what the test purports to assess.

Sattler (1982), and Salvia and Ysseldyke (1985) have outlined the criteria for judging an adequate test. Included is the requirement that a technical manual should accompany the test. Among the information the manual should outline is the standardization sample, the validity and reliability, the purpose of the test and the special qualifications required of administrators. It is interesting to note that Ebel (1979) reviewed 121 intelligence tests and found that the different constructs of intelligence resulted in different tests sampling different behaviors and producing different scores. Further, technical inadequacy (that is, inadequate standardization, lack of validity and a lack of reliability) is a frequently cited criticism (Lyon, 1983; Hilliard, 1979; and Williams, 1971). For example, the manuals for the 1972 edition of the Stanford-Binet, the Peabody Picture Vocabulary Tests Revised and the Cognitive Abilities Test offer no, or at best, meager data on the validity of these instruments (Salvia and Ysseldyke, 1985).

Another recurring criticism is that many of the major measures of intelligence are variations of the original Binet, using unchanged items that date to the beginning of the century. Zimmerman and Woo Sam (1973) argued that despite changes in society, children continue to develop in the same way, leading to the conclusion that such items still may be valid estimates of ability. However, Salvia and Ysseldyke (1985) in reviewing the 1972 edition of this test stated that the test has questionable merit resulting from an inadequate standardization description and a lack of validity and reliability data. Kennedy et al. (1961) found the ceiling was insufficient for bright adolescents despite the use of the Binet as a measure of giftedness.

It is not the intent of this section to review specific tests in detail, since that information is available in sources such as Sattler (1982) and Salvia and Ysseldyke (1985). However, a few summative comments regarding selected popular instruments follow as examples of the type of issues test users should consider. For example, while evidence for the reliability and validity of the Wechsler Scales is satisfactory, Salvia and Ysseldyke (1985) stressed the need for examiners to go beyond the more global verbal, performance and full-scale scores to look at individual performances on specific subtests.

The Peabody Picture Vocabulary Test - Revised, in one sense is an individual scholastic aptitude test. However, it is not a test of general intelligence. Instead, it measures only one facet of general intelligence - vocabulary. While adequate reliability data are available, there are no data on the validity of the measure. Used properly, and with the awareness that it samples only receptive vocabulary, Salvia and Ysseldyke (1985) concluded that the Peabody Picture Vocabulary Test - Revised can serve as a screening device.

With regard to whether a test should be used as a screening device as in the case of the Peabody Picture Vocabulary Test - Revised, or as a diagnostic device as in the case of the Wechsler Scales, the test user should consult the technical manual as well as associated reviews of the instrument. Determining the purposes for which the test was designed and the effectiveness of the test in meeting these purposes, is crucial.

Group administered intelligence tests have additional limitations which must be recognized by test users. For example, Salvia and Ysseldyke (1985) stated that group tests typically have a number of levels designed for use in specific grades. Further, students of the same age frequently are enrolled in different grades. Interpretation to compute mental ages is based on grade sampling results. At the extremes of the age or grade level for which the test was designed, the chance of error is intensified as a greater extrapolation is required.

Other limitations of group tests identified by Salvia and Ysseldyke (1985) include the fact that group tests are often standardized in schools which may exclude various types of students from their populations. An example of one such exclusion may be the severely disabled and the early school leaver. Further, as these samples typically include only volunteers, a sample bias is introduced. A final limitation regarding the standardization group rests in the fact that sample jurisdictions which are assumed to be representative of the total population are selected, rather than individuals who are known to represent a full range of the student types.

Despite the additional limitations of group tests, the Cognitive Abilities Test has been found by Salvia and Ysseldyke (1985) to be adequately standardized with good reliability. However, the evidence for validity is meager. The 1979 edition of the Otis-Lennon School Ability Test according to Salvia and Ysseldyke (1985), demonstrates the necessary technical adequacy to be used as a screening device.

Examples of these alternatives include work by Feuerstein (1979 and 1982) and his colleagues on the Learning Potential Assessment Device. The Learning Potential Assessment Device is a method assessing the potential of children, adolescents, and adults for growth in specific cognitive processes, first by guided exposure to problems and processes of thought and subsequently by their own independent efforts. The method is derived from, and contributes to the growth of, a theory of structural cognitive modifiability. The Learning Potential Assessment Device based on Feuerstein's Theory of Cognitive Modifiability, is characterized by an assessment strategy called dynamic assessment whose two distinguishing features are (a) assessment of fluid processes of thought, perception, learning, and problem solving rather than assessment of static faculties and/or the products of prior learning, and (b) carefully structured teaching of cognitive principles and processes followed by assessment of mastery and generalization of those principles. The goals of a learning potential assessment are: to identify well-developed cognitive functions; to identify deficient cognitive functions; to assess the response to teaching of cognitive principles and strategies; to estimate the kinds and amounts of investment required to overcome cognitive deficiencies; and to sensitize examiners and students to the processes involved in confronting and coping with a variety of tasks. It is important to note that the dynamic assessment of learning potential through the Device is not for everybody. For example, it is not indicated for persons who are successful with conventional tests and in conventional test situations. Also, the Device should not be used for classification purposes. Its supportive evidence is particularly suggestive of its usefulness with children from minority groups (Feuerstein, 1979). More research is required before a clear conclusion should be drawn. Further, Feuerstein acknowledges two disadvantages to the use of this test. It requires that the examiner possess special personality characteristics and training in the use of the technique. Also, the assessment procedure is lengthy resulting in an increased cost factor. Cummins (in press, p. 29) suggests that perhaps one of the major advantages of the Device is its "explicit acknowledgement of the central role played in the assessment process by the examiner's subjective intuition, skill and experience."

Another recent alternative has been the development and release of the Kaufman Assessment Battery for Children (1983). The Kaufman Assessment Battery for Children has as its primary goals:

- to measure intelligence from a strong theoretical and research basis
- to separate acquired knowledge from problem-solving ability
- to yield scores that translate to educational intervention
- to include novel tasks
- to be easy to administer and objective to score
- to respond to the needs of preschool, minority group, and exceptional children.

Kaufman and Kaufman (1983) wanted to produce a test that would answer much of the recent criticism of standardized intelligence tests: complaints that existing tests discriminate in favor of white, middle-class children; that tests are culturally biased and may unfairly label children; and, that the scores yielded by the tests do not readily translate into educational intervention.

Based on the simultaneous/successive model of processing information by Das (1984) and the work of Luria (1966), the Battery is concerned with the mental functioning involved in problem-solving. The Battery differs both from content-based intelligence tests and from tests employing a psycholinguistic model of communication. Content-based tests, such as the Revised Wechsler Intelligence Scale for Children, assess intelligence on the basis of language, problem-solving ability, and acquired knowledge. They measure a child's verbal or nonverbal response to the content of each test item. Tests based on a psycholinguistic model of communication, such as the Illinois Test of Psycholinguistic Abilities, primarily assess a child's verbal or motor response to auditory or visual stimuli. They are concerned with the nature of a stimulus and the response. The Kaufman Battery, in contrast, is concerned with what happens between the stimulus and the response and with how a person solves a problem mentally.

Both mental processing and achievement are assessed in the Battery. The Mental Processing Scales measure the student's ability to solve problems sequentially and simultaneously, with emphasis on the process used to produce correct solutions rather than on the specific content of the items. In contrast, the K-ABC Achievement Scale focuses on acquired facts and applied skills and therefore measures knowledge that a student has gained from the home and school environment.

The Kaufman Battery was standardized on 2,000 children, 100 at each half year of age between 2 years, 5 months, and 12 years, 5 months. The sample was stratified on the variables of sex, parental education, race or ethnic group (white, black, Hispanic, other), geographic region, community

size, and educational placement. The latter variable ensured that the sample would include proportional representation of exceptional children. Socio-cultural norms are also available.

A focal issue of the Journal of Special Education (Volume 18, Fall 1984) has been devoted to a review of the Kaufman Battery.

Another alternative in the measurement of intelligence, which has evoked a continuing interest over time is the Piagetian Tests developed by Jean Piaget (1952). Piaget's theory indicates that children at different stages of development show a great deal of variation in their mental organization, and this suggests that test content must vary with the cognitive stages of the child. Clarizio (1982) offers the reader a review of the more prominent Piagetian scales. Developmentally, the scales go from the infancy years through the middle childhood years. As of yet, there appears to be no commonly used battery for adolescents that is available outside of research settings. However, there is some suggestion that psychologists might find the Piagetian Tasks useful in their work with various exceptional groups such as the learning disabled, the mentally retarded and the gifted, as well as bilingual children (De Avila and Havassy, 1974; Eimers, 1971; LeRose, 1978; and Millar and Walker, 1981).

Despite the criticisms as well as the limitations and misuse of intelligence tests, no rival to the intelligence test has been demonstrated with regard to the general predictive ability of academic performance in school (Mussen et al., 1973).

Nevertheless, as stated by Sattler (1982), it is important for intelligence test users to note that:

1. An intelligence test score reflects the individual's interactions with the environment.

2. An intelligence test score can change in the course of development, even after the age of 6 years.

3. Intelligence test scores are only estimates.

4. Intelligence test scores do not measure the entire spectrum of abilities.

5. Intelligence test scores are typically specific to the test used and should never be assumed to be the same on a different test.

6. No intelligence test or battery of tests can provide a complete picture of an individual.

7. An intelligence test score represents a snapshot of the individual at a particular time.

A further concern for Canadians is that the more widely used and known intelligence tests are for the most part imported into Canada with the result that their standardization samples are rarely representative of Canadian people. However, if selected with care, with an understanding of their strengths, weaknesses and purposes and if used by professional, competent personnel, intelligence tests can be employed for the student's educational and personal/social benefit.

Examiner Variables

Lyon (1983) states that the quality of information that a testing process generates is directly related to the qualifications of the person selecting and administering the test. It is expected that the tester will be skilled and knowledgeable in establishing and maintaining rapport, in administering and scoring tests, and in analyzing results. A skilled examiner will not overlook important contextual variables such as the history and current life circumstance of the individual being tested as well as the examiner's personal and theoretical orientations. Indeed, the personality of the examiner may have an important effect on the examinee (Bersoff, 1973).

The test administrator must be aware of three critical areas where potential errors in judgment could occur. First, a technically inadequate test could be selected. Second, an adequate test could be used for the wrong purpose and, finally, there could be a failure to match the characteristics of the examinee with the tests. In practice, adequate testing devices have been administered incorrectly (Ross, DeYoung and Cohen, 1971; and Warren and Brown, 1973), and/or used unwisely and unfairly (Anastasi, 1961, 1967; Ebel, 1975; Laosa, 1977; Mercer, 1974; and Schmidt and Hunter, 1974). Systematic bias in the test scores of individuals may also be induced as a result of the examiner's expectancy (Mishra, 1980). The literature seems to suggest that a halo effect may occur in which examiners appear to have a tendency to overrate responses for subjects given high and positive expectancy (Cohen, 1965, and Donahue and Sattler, 1971). Graziano, Varca and Levy (1982) in their review of 29 studies which focused on the interaction between the examiner's race and the examinee's performance found that in some cases different races appear to systematically elicit different performance. While these studies on race and expectancy are inconclusive, they do suggest the possible impact of these factors in the testing process.

In summary, the examiner is engaged in an extremely complex two-person interaction. A linguistic, cultural and/or socio-economic gap, reflecting the different life experiences of the examiner and student, requires sensitivity in order to avoid a negative influence on the student's performance. It would seem that to avoid such problems requires a test administrator to be both "test wise" and "student sensitive".

Interpretation Variables

Adequate testing devices that are administered appropriately may still suffer from misinterpretation (Salvia and Ysseldyke, 1985) and/or over-interpretation (Bersoff, 1973). For example, research indicates that considerable skill is required in scoring ambiguous responses (Baughman, 1971 and Franklin et al., 1982). "Student sensitive" interpretation should reflect an awareness of the student's current life circumstances, his/her developmental history, and the conditions under which the student took the test. Often it is assumed that the student being tested has been exposed to comparable but not necessarily identical acculturation patterns relative to the standardization sample. Important differences frequently exist with respect to child rearing practices, expectations, aspirations, language experiences and availability of and involvement in informal and formal learning experiences. Without localized and/or pluralistic norms the chance becomes greater that, for these reasons, the interpretation of results may be less than meaningful should these factors not be considered by the examiners.

A good test interpretation takes into account not just the student and the test situation but also the resulting scores. The score from any one of a number of diverse intelligence tests is meaningful only if the person who is interpreting it understands how it was derived. Such a score is nothing more than a number if stripped of all relation to the test from which it was derived, of all relation to the individual to whom the test was administered, and of relation to the purpose for which it was given. A meaningful interpretation of a score cannot begin to be achieved if the source of the particular test score is not known, if the nature of the test and how it was administered are ignored, and/or if the purpose of giving the test has not been determined. Blackmore (1984, page 23) suggests that a procedure be followed in dealing with the interpretation of any test. As an example, the administrator should find out the nature of the test used to generate the score, how it was administered, the purpose of the test, and the distribution of test scores (including the mean, standard deviation and standard error of the test). In interpreting test results, Hilliard (1979) states that an aggregation of numbers (test scores) should not be confused with the aggregation of comparable units of mental behavior or skill. For example, on the Wechsler Scales the score from the block design is aggregated with that of vocabulary to give an overall score. However, aspects of mental behavior are not additive in this simple sense.

Having focused on the instruments, examiner variables and interpretation variables, three issues remain. One issue is concerned with the rights of the student. A second issue concerns the use of intelligence test scores in educational decisions, and the third issue concerns those groups that may require special consideration in the administration and interpretation of intelligence tests.

THE RIGHTS OF THE STUDENT

Focus on the student is vital, for it is the student who is directly and significantly affected by choices in the usage and interpretation of an intelligence test as well as in the resulting educational decision making.

The Canadian Charter of Rights and Freedoms presents an opportunity for schools and specialists, such as psychologists, to re-examine the rights of students in intelligence testing and the use of test results. While the Charter does not specifically reference children's rights, questions regarding the degree and in what circumstance the terms, "everyone", "every citizen", "any person" and "every individual", apply to a child or a student, need to be answered. The intent is to provide basic human or individual rights such as freedom of speech, natural justice, right to privacy, right to dissent, and the right to equal opportunity "to such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society" (Canadian Charter of Rights and Freedoms, Section I).

Damon (1977) suggested that where appropriate, the individual who is being tested have access not only to the testing results but also to clear, concise answers to the following questions:

- Why is the test being given?
- What test is being given and what are its limitations?
- When and where will the test be given and by whom?
- Who will receive the test results? How will they be reported? Where will they be stored?
- When will the test results be destroyed?
- How will the results be used, by whom, and with what degree of confidentiality?
- What are the consequences if a student does not take the test?
- Will students be able to see the test beforehand?
- How will students receive the test results?

Intelligence test scores must not only be interpreted properly but must also be communicated effectively. A suggestion found repeatedly in literature is the need for a reporting system which will have greater educational value. One suggestion to improve the usefulness of reports

lies in avoiding a discussion of the results in terms of psychometric constructs, but rather to discuss the results in operational terms or in terms of behaviors (Turnbull, 1979). It would seem appropriate that since the testing is done for educational reasons, the report should reflect appropriate educational interventions based on those data.

Since judgements are not infallible, an appeal process ought to be available (Bergin, 1984). Intellectual assessment results, and decisions based on such assessments, are no exception. Section 27 of the Canadian Charter of Rights and Freedoms indicates that schools and school personnel will need increased sensitivity to penalties and/or loss of learning opportunities inflicted on students because they speak languages other than French or English as their mother tongue. Testing procedures require examination and change if, in fact, they act to penalize such students. The challenge to educators according to Mackay (1983) is to use the Canadian Charter as "a catalyst" to reasonably re-examine the use to which intelligence tests and their results are put.

IMPLICATIONS OF INTELLIGENCE TEST RESULTS IN DECISION MAKING

Labelling. Every description, in some sense, implies a label. When the term "normal range of functioning" is used to describe a child, the child is being labelled. However, in terms of expediting discussions with fellow professionals, labels do have a certain value. Labels act as a guide and a convenient device for communicating a set of concepts. However, labels should be thought of only as a point of reference, not as a means of denying students an opportunity to learn or to enter into a program. Sattler (1982) stated that educational decisions should be guided by a student's performance, not by a classification system with arbitrary cutoff points.

Intelligence tests measure abilities and skills acquired by children with highly differentiated exposure to opportunities at a period in their development. As well, intelligence tests provide profiles of current functioning (Salvia and Ysseldyke, 1985). Despite these facts, resulting labels often interfere with a clear perception of the individual (Rosenham, 1975). Those judged to be "below average" are not likely to receive, in most schools, the kinds of educative opportunities available for children judged "above average". Children placed in "remedial" or other special education programs and in lower-level tracks are often viewed as failures and this may result in lowered expectations in some cases. Students in such settings quickly learn to view themselves as failures who produce little (Perrone, 1977). Rose and Rose (1980) state that the gravest problem with intelligence tests is that the labels resulting from scores on the tests do little to promote teaching.

Also, labels are often misunderstood by students and parents and are regarded as inaccurate and humiliating by those who are so classified

(MacMillan, 1977). Dirks, Bushkuhl and Marzanos' (1983) findings support the idea that labelling by means of an intelligence test score has a potentially negative impact on parents and their children. Before the actual testing occurs, examiners should emphasize to parents that "real-life" achievement is dependent upon a variety of factors in addition to an intelligence test score.

Indeed, there has been some suggestion that the term intelligence is problematic and contributes to the labelling process. The term is linked to the belief that intelligence is unitary, fixed and predetermined (Garcia, 1981). Reschly (1979) suggested that while accurate communication of test results and the meaning of test scores is universally endorsed by psychologists, the term itself is a significant barrier. The often suggested solution to these concerns is to abandon the term and replace it with other descriptors such as school ability or academic aptitude (Lennon, 1980). However, these terms also have definitional problems.

Placement. Smith and Knoff (1981) found that intelligence scores had a powerful, paralyzing effect on flexibility in problem-solving as it pertained to programming decisions for both school psychology and special education graduates. For example, intelligence test scores carried more weight than did adaptive-behavior skills. The intelligence test score was the most salient factor in the decision making process and tended to lower placement opinions. Supportive to this is the work of Berk, Bridges and Shih (1981) who found that intelligence level was a critical factor in recommended placement in classes for the mentally handicapped. Mercer's (1973) findings, though controversial, also are supportive of the critical role intelligence test scores play in placement.

SPECIAL POPULATIONS

The consequences of intelligence test use have been of particular concern to various special interest groups. As Salvia and Ysseldyke (1985) and others have very carefully pointed out, these consequences are the basis of much of the critical discussion surrounding intelligence testing. The following section briefly reviews some of the more salient issues.

Minority Students. Through the 1970's, the fairness of intelligence tests for minority students became a central issue. The use of these tests in the referral-placement process came to be viewed as part of an overall pattern of "institutional racism" (Jones and Wilderson, 1976). The tests have been held responsible for classifications deemed to be humiliating and stigmatizing to the individual, racist with respect to groups, and, most critically, irrelevant and damaging to educational programming.

Reynolds (1982) divided the criticisms into six basic areas:

1. The content of the tests is unfamiliar to and inappropriate for use with minority children.

2. Standardized test samples do not reflect all groups; minority students are not adequately represented because their numbers are too small to have any impact on item selection.

3. Examiner and language bias is present since most psychologists are white and speak only standard English, which intimidates and confuses minority children.

4. Inequitable social consequences result when minorities, already at an economic and educational disadvantage, are relegated to inferior programs because of test performance.

5. The tests measure different attributes when used with children outside of the mainstream white middle-class culture.

6. Issues such as the educational and psychological development of economically disadvantaged minorities are ignored.

While the question of cultural bias is a long-standing one which can be traced to the historical roots of intelligence tests, most current practitioners agree that intelligence tests do not measure "innate ability" but "academic potential" or the ability to perform in a school setting (Laosa, 1977 and Kamin, 1974). Also recognized is the fact that tests must be used judiciously with students from backgrounds other than those of the dominant cultural group because the potential cultural or linguistic bias could result in inappropriate educational decisions.

English-As-A-Second-Language (ESL). A survey of Toronto school districts by Desosaran (1976) found both a disproportionately high number of low socio-economic status children and immigrant children, who had learned English as a second language, in special education classes in elementary schools and in vocational streams at the secondary level. However, socio-economic status seemed to have less weight in placement decisions than did the learning of English as a second language. To combat these kinds of findings, a few Toronto Boards have adopted a policy of delaying the administration of formal diagnostic tests and group tests of ability and achievement until students have been in Canada at least two years. Samuda (1980) and Samuda and Crawford (1980) indicated, however, that this change had not gone beyond the Toronto boundaries.

In a study of 428 students referred for assessment, Cummins (1980) found that children born outside of Canada (54% of the sample) were referred for language (68%) and attendance (63%). One of the most common

concerns identified by teachers regarded the evaluation of the students' true academic potential. Psychologists typically administered one of the Wechsler Scales. Of the sample, students born in Canada performed consistently better than those born abroad on both the intelligence and achievement tests. The study documents the following knowledge/practice gaps:

- Psychological tests access student's present academic functioning not potential.
- The Wechsler Scales subtest pattern of minority children, which may provide diagnostic clues, is often not used.
- ESL immigrant children take, on the average, between five to seven years to approach grade norms in English cognitive/academic skills.
- Interpersonal communication skills are very different from cognitive/academic language proficiency.
- The ESL children's academic difficulties are not caused by the use of non-English languages in the home.

Cummins suggested that in the development of policy regarding minority ESL children and testing/placement decisions, the following questions may have some relevance:

- When teachers monitor a student's rate of learning after initial placement, what criteria are used to determine whether or not the student's progress is satisfactory?
- What stages do students of different ages go through in acquiring English proficiency, and how long are academic difficulties likely to persist?
- What is the basis for delaying psychological testing for two years as opposed to one, three, or four years?
- What criteria are used to distinguish between reading difficulties which arise because English is a second language and those that reflect specific learning disabilities which may require specialized remediation procedures?
- To what extent does the delay of formal diagnostic procedures deny minority language children the possibility of early identification and remediation of specific learning disabilities?

Native Children. An extensive examination of the results of validation studies focusing on the use of cognitive tests in the psychoeducational assessment of Native children has been reported by Matheson (1983). He stated that intelligence tests are not necessarily equally valid across Native populations for any given purpose as their correlations with school achievement variables are not totally consistent. Further, children who are sampled in the Native communities tend to obtain low scores on tests which require extended answers in English, although they may obtain scores at or above the national average on tasks with similar content but different response modes. Notwithstanding the above point, the Verbal-Nonverbal distinction is considered a simplistic description or explanation of Native children's performances on cognitive tests. As an example, group profiles on sets of nonverbal tasks have shown considerable scatter (Wittshire & Gray, 1969). Discrepancies among Inuit, Indian, and white sample factor patterns for batteries of academic, linguistic, and/or spatial tests, plus Feldman and Bock's (1970) findings regarding the rank order of difficulty levels, suggest that in the patterns of cognitive test scores of Native children, there are factors other than verbal intelligence or spatial ability which cause different scores.

Recognizing the concerns raised by Matheson (1983), there is an increased probability that the use of intelligence tests with this population will result in inappropriate educational placement. The Task Force on Crosscultural Assessment of New Jersey (1980) strongly suggested that a moratorium be placed on the use of standardized intelligence tests in the assessment of black or native Americans because such assessment could result in inappropriate educational placement for these children. The Task Force on Crosscultural (1980) Assessment made the following suggestions:

- For each child from a linguistic minority, a multidisciplinary assessment team should be composed. This team should include at least one person who speaks the child's language and is familiar with the child's culture and one person experienced in bilingual education, preferably in the child's language.

- A determination of the level of communication skills and dominance in both English and the native language should precede all other assessment procedures.

- The child should be observed by the assessment team in a variety of settings including the classroom. An attempt should be made to describe how the child functions in these settings.

- A team member thoroughly knowledgeable with the child's language and culture should prepare a home survey based on a visit to the child's home. This team member would attempt to discover what language the family normally speaks, what language is spoken in the community, what exposure the child has had to the English-speaking core culture, and what previous education the child has

had. The staff member would gather as much information as possible about the child's previous history and experience.

- The assessment should focus on determining how the child functions, socially and academically, in both English and the native language.

- All assessment techniques and procedures, including standardized tests, should be appropriate to the child's age, culture or language, socio-economic background, communication skill level in English, familiarity with the core culture and ability to function independently in a variety of social settings. All procedures and techniques should be administered by an appropriately qualified professional who is thoroughly familiar with the child's culture and language.

Learning Disabled. It is often assumed that since learning disabilities are characterized by discrepancies in cognitive functioning, these variations should be reflected in a test of cognitive skills or an intelligence test. Typically the assumption is more specific in that examiners seek out profiles in the results, particularly in the Wechsler Scales, that correspond to known learning disabilities. However, one does not find a basis for such a profile in the empirical research. While a few studies have reported positive findings of a characteristic profile, the profiles reported have been contradictory and lack replication (Millar and Walker, 1981). This contradiction is, in part, a reflection of the heterogeneous nature of the category of learning disabilities.

Patterns sometimes thought to be representative of learning disabled students are the result of either a substantial discrepancy between Verbal-Performance portions of the Wechsler Scales or a great amount of scatter among subtests over the whole test. These patterns were in part derived from a study by Clements and Peters (1962). However, this study used minimal brain dysfunction individuals as subjects. Kaufman (1979a, b) cautions that these patterns must be compared to those of typical children to ensure that the same fluctuations do not occur with similar frequency in normal populations.

Behaviorally Disordered. Seriously disturbed children often function poorly on tests. Even tests like the Stanford-Binet and the Wechsler Intelligence Scale for Children - Revised often underrate the emotionally disordered child's level of intellectual functioning (Haring and Phillips, 1962). Jensen (1980) stated that fluctuations of as much as 20 to 40 IQ points are not uncommon in such a population. Factors in the student's behavior which could adversely affect test performance should be noted by the examiner.

Physically Disabled. In general, three approaches are taken when using intelligence tests with exceptional students who have difficulty responding to traditional devices. Examples of exceptional students in this category are the blind, the deaf, the multiple sensory handicapped and children with cerebral palsy. Test items may be adapted. However, this practice overlooks the fact that the test has been standardized using a different standardization group. Another option is to use tests that require different modes of responding. Unfortunately, such tests are standardized on non-handicapped populations. Tests designed for and standardized on handicapped populations may be used but a limited number of such devices are available. Further, Salvia and Ysseldyke (1985) have noted the limited validity of such tests. In summation, the test administrator must be sure that the response requirements are both fair and reasonable and must be cautious in the use of norms.

Gifted. Over-reliance on IQ test scores to determine giftedness has been cited as problematic (Kirschenbaum, 1983 and Sternbery, 1982). Sapon-Shevin (1984) stated that this over-reliance leads to an under-representation of minority students and a disproportionate over-representation of upper and middle-class white students who are identified as gifted. According to French (1974) and Martinson (1974) the two most frequently used instruments for the identification of intellectually gifted children appear to be the Stanford-Binet and the Wechsler Scales. Despite the excellent reliability and validity of the Wechsler Scales, Harrington (1983) outlined ten basic problems in using the Wechsler Scales and other tests of intelligence to identify the gifted and talented:

1. The use of arbitrary "cutoff scores". When the standard error of measure is considered, there is no statistical difference between a student's score of 126 and a score of 130 on the Wechsler Scales. Nevertheless, many school jurisdictions use a cutoff point of 130.

2. Specific issues in specific tests. Both the Stanford-Binet and Wechsler Scales have relatively low ceilings (Sattler, 1982). Further, Gallagher (1966) indicated that the Stanford-Binet has a higher IQ range than the Wechsler. Therefore, test results may reflect an underestimation of the student's abilities because the scales may not be challenging enough for the student at the upper extreme of the test.

3. There is a lack of interchangeability of test results. Although there is a high positive correlation between the Stanford-Binet and Wechsler Scales, a perfect, positive correlation between the scores rendered by the tests does not exist. Beeman's work (1960) indicated that there may be a difference of 10 points in favour of the Binet. This variance in test results, in a case of reassessment on two different tests, may affect a student's eligibility for a gifted program.

4. The impact of practice effects. When the same instrument is used in a reassessment, a possibility exists that a practice effect or test wise behavior may inflate scores, especially on time and performance type items.

5. Early identification issues. Some educators suggest that identification might occur as early as kindergarten with individual intelligence tests. This statement fails to take into account the developmental nature of the child and, by extension, the capacity of intellectual behavior to grow (Sattler, 1982).

6. The potential exclusion of minority children. When the identification process consists predominantly of standardized intelligence tests, those minority students who have not participated in the dominant culture may be under-identified (Douglass, 1969, Passout, 1966; and Renzulli, 1973).

7. The impact of stress. To a potentially gifted student, the desire to perform well on an intelligence test, which may qualify the student for a gifted program, may produce very great stress and may actually interfere with test performance (Freehill and McDonald, 1981; and Sternberg, 1982).

8. The potential exclusion of "low incidence" populations. The hearing impaired, the visually impaired, and the orthopedically impaired remain unrepresented or under-represented in the standardized norms of the Binet or Wechsler Scales. In addition, such disabilities may interfere with test performance.

9. The specific problems with group tests. Pegnato and Birch (1959) state that group intelligence tests under-refer intellectually gifted students. The higher the level of ability the greater the discrepancy between a student's group intelligence score and individual intelligence score, resulting in a spread of as much as 30 points at the upper levels of intellectual ability. This effect occurs at the upper limits of the group test because students may be required to perform almost perfectly on all items to be identified as gifted. However, Sattler (1982) suggests that group intelligence tests are successful about 45% of the time in identifying gifted children.

10. Specific processing styles may be over-represented. The design of tests and scoring criteria are such that convergent type responses are rewarded while divergent type responses often receive no credit.

Martinson (1974) suggests that an identification procedure which uses multiple criteria including individual intelligence tests, teacher observation and nomination, previous accomplishments, group achievement scores, creativity test scores and group intelligence test scores, would be more helpful in identifying the gifted and talented.

Mentally Handicapped. Berk, Bridges and Shih (1981) state that intellectual level is one of the two most powerful reasons for the recommendation regarding placement into Educable or Trainable Mentally Handicapped classes. This finding suggests that intelligence scores figure significantly in placement decisions. A study by Smith and Knoff (1981) is supportive of Berk, Bridges and Shih's findings, with intelligence test scores considered a more substantial predictor of special class placement than either adaptive behavior or academic information.

In an effort to grapple with the relationship between IQ and later adjustment in life, commentators such as Garfield and Wittson (1960a, 1960b), Mercer (1975), and Zigler (1967) have moved to include social competence as a critical variable in defining mental handicap. This inclusion recognizes that the behaviors comprising intelligence test scores represent only a very few of the many variables accounting for later competence. Reflecting these comments, a measure of adaptive behavior may be more useful in assessing possible members of this population than is an intelligence test. As part of a comprehensive multi-dimensional assessment, this method may offer diagnostic information important for program planning.

SUMMARY

At the heart of an intellectual assessment is the continuing and expanding effort to ensure that assessment procedures result in positive benefits for students. The emphasis must be on the process rather than on a set of instruments (Reschly, 1979).

A review of the literature revealed several suggestions for guidelines to change assessment procedures in a manner which would promote unbiased testing. Basic to any procedure is that intelligence tests be given for a specifically stated purpose that is agreed upon in advance by those who are to use the results. This is one way to ensure that the information derived from intelligence tests will be helpful in making better decisions. Further, intelligence tests should be viewed as only one source of data for such decision making.

While intelligence tests are not without technical problems, it would appear that, like most tools, the skill of the user and the use to which tests are put are equally important factors to be considered by school personnel. The professional educator needs to be critically aware, "test wise", "student sensitive" and sensitive to both the issues and alternative procedures in order to promote an educationally useful intellectual assessment.

SUMMARY OF THE SURVEY OF ALBERTA SCHOOL JURISDICTIONS

When Alberta school jurisdictions were surveyed in 1984, to determine the status of intelligence test use in Alberta schools, responses were received from 101 of the 147 jurisdictions. Those responses indicated that slightly less than half had written policy or guidelines for the use of intelligence tests within the jurisdiction. Special provisions for students whose mother tonque is not English, ethnic minority students, and gifted and talented students were provided by very few jurisdictions.

The most frequently reported individual intelligence measure used by school jurisdictions was the WISC-R. The Peabody Picture Vocabulary Test - Revised was cited as the second most frequently used individual test. The Wechsler Preschool and Primary Scale of Intelligence and the Standord-Binet were also popular, but less so than the first two. The Canadian Lorge Thorndike was cited as the most frequently used group intelligence measure. The Otis Lennon School Abilities Test and the Cognitive Abilities Test were chosen second and third respectively. School jurisdictions responding cited the following four criteria as most critical for test selection: appropriateness of the test for the needs of the student, the purpose of testing, the referral problem, and adequate test validity and reliability.

Approximately two-thirds of the responding school jurisdictions indicated that a student should be given an individual intelligence test only upon referral. Reflecting this, individual intelligence tests were most often given to determine program needs or as a step in the referral process. It was indicated that slightly more than one-half of the test administrators had related courses beyond those required by the Alberta Education Learning Disabilities Fund Registry.

Group intelligence tests were most frequently used as part of a general screening procedure. However, there was no consistent pattern as to when or how frequently school jurisdictions used group intelligence tests. Most frequently, the tests were reported as administered by classroom teachers. Some jurisdictions indicated having stopped using group intelligence tests for general screening purposes.

STATUS OF POLICY ON INTELLIGENCE TESTS IN CANADA

The results of a survey done in 1986 indicate that no Canadian province has a comprehensive written policy regarding the use of intelligence tests. The following list presents each province's position or initiatives in this area.

- Newfoundland has guidelines on program planning which address assessment. In general, the Peabody Picture Vocabulary Test-Revised, the Raven's Progressive Matrices and the Slosson Intelligence Tests are used as screening mechanisms and the Wechsler Intelligence Scale for Children-Revised (WISC-R) is used for placement. Usually, the WISC-R is used in combination with a test of functional ability.

- Prince Edward Island reports that decisions concerning intelligence tests are left to local school boards. This approach translates into counsellor/principal decisions. There is no intent to change this approach. There has been a reaction against the use of intelligence tests with the result that usage has been discouraged. Tests are done on a referral basis (only the WISC-R and WAIS are used). Group tests are highly unpopular due to critical reviews on group tests in current research literature.

- Nova Scotia school personnel do testing as they think necessary for the proper placement and remediation of children with school problems. Jurisdictions decide when and under what circumstances testing occurs.

- New Brunswick school jurisdictions set policies and guidelines regarding the use of intelligence tests. There is a trend toward the use of ranges in reporting rather than the dissemination of scores.

- Ontario reports that such decisions are left up to the jurisdiction. In general, the intelligence test is used as one of several instruments. Policy exists regarding psychological testing and assessment of pupils, which requires individual psychological tests (intelligence tests) to be given only if there is written parental consent and if it is administered by a registered psychologist.

- Manitoba has no set procedures which all users of IQ tests must follow. School jurisdictions are responsible for carrying out such assessments where necessary and appropriate.

- The Saskatchewan Department of Education has, for the last seven years, focused on identification, assessment and screening procedures. There are statements in place on test protocol and times for administration, particularly as applied to exceptional children. These statements focus on assessment in general and not on intelligence tests in particular.

- British Columbia officially discontinued the required use of intelligence tests in 1973. Intelligence test scores no longer appear on a student's permanent record. School jurisdictions may use intelligence tests if they choose, however. In practice, some school jurisdictions use intelligence tests and some do not. The WISC-R is the commonly used test in those cases in which intelligence testing is used.

Common threads emerged from the reports from other Canadian provinces. They were:

1. A concern for parents and their inclusion in educational decision making;

2. The need to ensure natural justice;

3. Decisions regarding which tests to use and when to test were considered to be largely professional ones to be made by school personnel and/or psychologists; and,

4. A shift in emphasis from placement to program decision.

POLICY OF INTELLIGENCE TESTS IN THE UNITED STATES

In the United States, federal legislation (PL 94-142) has, in part, responded to the criticisms of testing and the lack of adherence to professional testing standards. This law legislates areas which must be incorporated within a state educational plan. United States Public Law 94-142 ensures the appropriateness of tests and testing procedures. Further, it demands accurate, reliable and meaningful evaluation of a child in determining the presence or absence of a handicapping condition. Services for gifted children are not included in Public Law 94-142 (Sapon-Shevin, 1984).

To be more specific, tests or testing practices must have an established validity. They must include tests and testing practices designed to assess areas of educational need, not merely those which are designed to provide a single general intelligence quotient. Tests must be administered by trained personnel in conformance with the instructions provided by the test producers and must be validated for the specific purpose for which they are being used. No single criterion is to be used for determining an appropriate educational program for a child. All evaluations are to be made by a multidisciplinary team, and the law indicates that the team must include a teacher or other specialist in the area of suspected disabilities. The law prohibits the use of assessment practices that penalize the child because of cultural or ethnic bias. Finally, tests whose stimulus and response demands exceed the physical capabilities of the child are viewed as inappropriate: the scores resulting from such tests might be inaccurate and not representative of the child's true capabilities.

Partially in response to this federal law, but more as a response for litigation (Larry P. versus Riles; PASE versus Hannon; to name only a few), some states (California, for example) have banned the use of intelligence tests. Nonetheless, it is often suggested that even in those states intelligence tests are still unofficially being used.

SUMMARY OF FINDINGS

The present literature regarding this subject, is inconclusive and no general consensus exists on a single operational definition of intelligence. Where consensus does exist, it revolves around three points:

* The nature of intelligence is culturally relative.

* Ability alone rarely determines performance.

* Intelligence is a dynamic, fluid and multi-trait construct.

In spite of this lack of general consensus, there is a degree of agreement that when properly used, intelligence tests do provide useful information which can be the basis for constructive planning. The key here is ensuring that the intelligence tests are selected carefully, administered properly, understood thoroughly and used judiciously.

The survey of Alberta school jurisdictions shows a lack of consistency, continuity and any uniform procedures across the province. Less than one quarter of all of the school jurisdictions in Alberta have guidelines regarding the use of intelligence tests. This situation is not unique to Alberta, however, as the Canadian survey revealed that no Canadian province has a comprehensive policy governing this issue.

In the United States of America, the use of intelligence tests is controlled by Public Law 94-142. As a result of this stringent legislation some States have completely banned the use of intelligence tests, probably to avoid litigation more than as the result of perceived lack of usefulness.

GLOSSARY

Criterion Referenced Tests - These measure a person's skill level in terms of absolute mastery of a particular subject or area.

Cultural Bias - When a test, which is normed on a particular cultural group, is used with individuals who are part of a very different ethnic and cultural background, these latter individuals may achieve rather poorly due to a lack of understanding of the terminology used.

Halo Effect - The tendency of an examiner to inflate the scores obtained, by departing from designated procedures, giving extra clues and reading things into a child's responses which are not actually there.

Intelligence Tests - (a) Individual - This type is designed to be administered to only one individual at a time. The questions and demands are usually given orally and the tester is able to observe the responses directly and thus gather quantitative and qualitative information.

- (b) Group - This type can be given to one person or several simultaneously. After the fourth grade, the directions are usually written rather than oral, and qualitative information is very difficult to obtain.

- (c) Culture Fair - This type is designed to measure general mental ability and to reduce, as much as possible, the influence of verbal fluency, culture climate and educational level.

Minimal Brain Dysfunction - The result of actual damage to the brain, which may have been caused by a head injury, anoxia, high fever, convulsion, tumor or lesion. This damage is manifested by a reduction in, or loss of, certain intellectual, social, emotional or physical skills.

Nonsystematic Observation - Watching for significant and revealing behaviors of the individual during a testing session.

Norms - An individual's performace is evaluated in terms of a comparison with other people's performances. The adequacy of these norms is essential for test accuracy and relies upon the representativeness of the norm sample, the number of cases in the norm sample, and the relevance in terms of the purpose of testing.

Norm Referenced Tests - These tests compare an individual's performace to that of his/her peers.

Power Tests - Timed tests which measure the student's total knowledge of the area being tested.

Quantitative Data - The actual scores achieved on a test.

Qualitative Data - Obtained through nonsystematic observations made while a child is being tested, and which tells us how he/she arrived at the answers given.

Receptive Language Vocabulary - The degree of understanding which a peron has with regard to the words which he/she is either reading or using in written and oral expression.

Reliability - The tendency of a test to provide consistently, similar scores in subsequent uses with the same person. If all other variables remain relatively constant, we would not expect a person's scores on repeated testings to fluctuate randomly if the test has good reliability.

Sample Bias - This occurs when procedures are introduced which limit the scope of the representative population; e.g., using volunteers rather than random selection of subjects.

Sampling - When a representative subset of the population is used to estimate the characteristics of the larger population, and to make inferences about this total population, then sampling is said to have occurred.

Speed Tests - Timed tests which measure a number of other factors in addition to knowledge of the area being tested.

Standardization - A test is standardized by comparing its performance to the performance of a group of subjects of known characteristics such as age, sex, grade level, etc., commonly called a normative group.

Validity - The extent to which a test measures that which its authors or users claim it measures.

- (a) Content Validity - The appropriateness of the test items, the completeness of the item sample and the level of mastery at which the content is assessed.

- (b) Criterion Validity - The extent to which a person's score on a criterion measure can be estimated from that person's test score.

- (c) Construct Validity - Directional hypotheses are formed on the basis of construct definition, construct theory and empirical research. These hypotheses are then tested and if the test results confirm the hypotheses, construct validity is said to exist.

REFERENCES

Anastasi, A. Psychological Testing. New York: MacMillan, 1961.

Anastasi, A. Psychological Testing. (3rd. ed.) New York: MacMillan, 1967.

Baughman, E. Black Americans. New York: Academic Press, 1971.

Beeman, G.A. A Comparative Study of the WISC and Stanford-Binet With a Group of More Able and Gifted 7 to 11-Year-old Children. Presentation to the California Association of School Psychologists, 1960, mimeographed.

Bergin, J.J. "Rethink Your Treatment of Students." The Canadian School Executive, 1984, 3:2, 13-15.

Berk, R.A., Bridges, W.P., and Shih, A. "Does IQ Really Matter? A Study of the Use of IQ Scores for the Tracking of the Mentally Retarded." American Sociological Review, 1981, 46, 58-71.

Bersoff, D.N. "Silk Purses into Sows' Ears: The Decline of Psychological Testing and A Suggestion for its Redemption." American Psychologist, 1973, 28, 892-899.

Blackmore, D.E. "Interpreting Test Scores: The Basics Revisited." The School Guidance Worker, 1984, 39:4, 17-25.

Bruner, J.S. Studies in Cognitive Growth. New York: John Wiley, 1966.

Clarizio, H.F. "Piagetian Assessment Measures Revisited: Issues and Applications." Psychology in the Schools, 1982, A, 421-430.

Cleary, T.A., Humphrey, C., Kendrick, A., and Wesman, A. "Educational Uses of Tests with Disadvantaged Students." American Psychologist, 1975, 30, 15-41.

Clements, S.D., and Peters, J.E. "Minimal Brain Dysfunctions in the School-Aged Child." Archives of General Psychiatry, 1962, 6, 185-197.

Cohen, L.S. An Experimental Manipulation of "Halo Effect". A Study of Teacher Bias. Unpublished manuscript, Stanford University, 1965.

Cronbachil. "Five Decades of Public Controversy over Mental Testing." American Psychologist, 1975, 30, 1-14.

Cummins, J. Psychological Assessment of Minority Language Students. Ontario Institute for Studies in Education, 1980.

Cummins, J. Assessment of Bilingual Exceptional Students. In J. Cummins, Special Education and Minority Students. (In press.)

Damon, J.P. "Questions You Should Ask About Your Testing Program." In P. Houts (ed.), The Myth of Measurability. New York: Hart Publishing Company, Inc., 1977.

Das, J.P. "Simultaneous and Successive Processes and KABC." Journal of Special Education, 1984, 18, 229 - 239.

Davis, F. Standards for Educational and Psychological Tests. Washington, DC: American Psychological Association, 1974.

De Avila, E., and Havassy, B. "The Testing of Minority Children - A New Piagetian Approach." Today's Education, 1974, 63, 72,75.

De Lisi, R. Intelligence, Intelligence Testing and School Practices. Revised Edition. ERIC Clearinghouse on Tests, Measurement and Evaluation, Princeton, NJ: December, 1980.

Desosaran, R. The 1975 Every Student Survey. Research Report #140, Toronto Board of Education, 1976.

Dirks, J., Bushkuhl, J., and Marzano, P. "Parents' Reactions to Finding Out That Their Children Have Average or Above Average IQ Scores." Journal of School Psychology, 1983, 21, 23-30.

Donahue, D., and Sattler, J.M. "Personality Variables Affecting WAIS Scores." Journal of Consulting and Clinical Psychology, 1971, 36, 441.

Douglass, J.H. Strategies for Maximizing the Development of Talent Among the Urban Disadvantaged. Paper Presented at the Annual Meeting of the Council for Exceptional Children. Denver, April, 1969.

Ebel, R.L. "Educational Tests: Valid? Biased? Useful?" Phi Delta Kappan, 1975, 57 (October), 83-83.

Ebel, R.L. "Intelligence: A Skeptical View." Journal of Research and Development in Education, 1979, 12:2, 14-21.

Eells, K., Davis, A., Havighurst, R.J., Herrick, V.E., and Tyler, R.W., Intelligence and Cultural Differences, University Party Press, Chicago, 1954.

Eimers, S. Assessment of the Concept of Conservation Among Mexican-American and Caucasian First and Second Grade Children. Masters Thesis, California State College, Long Beach, 1971.

Feldman, C.F., and Bock, R.D. Cognitive Studies Among Residents of Wainwright Village, Alaska. Arctic Anthropology, 1970, 7 (1), 101-108.

Ferguson, G.A. "On Learning and Human Ability." Canadian Journal of Psychology, 1954, 8, 95-112.

Feuerstein, R. The Dynamic Assessment of Retarded Performers: The Learning Potential Assessment Device, Theory, Instruments and Techniques. Baltimore, MD: University Part Press, 1979.

Feuerstein, R. The Learning Potential Assessment Device, A Manual. Unpublished, 1982.

Franklin, M.R. Jr., Stillman, P.L., Burpeace, M.C., and Sabers, D.C. "Examiner Error in Intelligence Testing: Are You a Source?" Psychology in the Schools, 1982, 19, 563-569.

Freehill, M., and McDonald, J. "Zeal: Essential to Superior Intellectual Achievements?" Gifted Child Quarterly, 1981, 25, 123-127.

Freire, P. Education for Critical Consciousness. New York: Seabury, 1973.

French, J.L. "The Gifted In M.V. Wisland (ed.)." Psychoeducational Diagnosis of Exceptional Children. Springfield, Illinois: Charles C. Thomas, 1974.

Fuller, R. In Search of the IQ Correlation: A Scientific Whodunit. Stony Brook, New York: Ball Stick Bird, 1978.

Gage, N.L., and Berliner, D.C. Educational Psychology. 2nd ed. Chicago: Rand McNally, 1979.

Gallagher, J.J. Research Summary on Gifted Child Education. Springfield, Illinois: Department of Public Instruction, 1966.

Gardner, H. Frames of Mind: The Theory of Multiple Intelligence. New York: Basic Books, 1983.

Garcia, J. "The Logic and Limits of Mental Aptitude Testing." American Psychologist, 1981, 36, 1172-1180.

Garfield, S.L., and Wittson, C.L. Comments on Dr. Cantor's Remarks." American Journal of Mental Deficiency, 1960, 64, 957-959(a).

Garfield, S.L., and Wittson, C.L. "Some Reactions to the Revised Manual on Terminology and Classification in Mental Retardation." American Journal of Mental Deficiency, 1960, 64, 951-953. (b).

Glaser, R. "The Process of Intelligence and Education." In L. Renicked, the Nature of Intelligence. Hillsdale, N.J.: Lawrence Emlbaum Associates, 1976.

Goodnow, J.J. "Concepts of Intelligence and its Development." In N. Warren (ed.), Studies in Cross-Cultural Psychology. London: Pergamon Press, 1979.

Gould, S.J. The Mismeasure of Man. New York: W.W. Norton and Company, Inc., 1981.

Graziano, W.G., Varca, P.E., and Levy, J.C. "Race of Examiner Effects and the Validity of Intelligence Tests." Review of Educational Research, 1982, 52:4, 469-497.

Guilford, J.P. The Structure of the Intellect. Psychological Bulletin, 1956, 53, 267-293.

Haring, N.G. and Phillips, E.L. Educating Emotionally Disturbed Children. New York: McGraw-Hill Book Company, Inc., 1962.

Harrington, R.G. "Caution: Standardized Testing May Be Hazardous to the Educational Programs of Intellectually Gifted Children." Education, 1983, 103:2, 112-117.

Hebb, D.O. The Organization of Behavior. New York: John Wiley, 1948.

Hilliard, A.G. The Ideology of Intelligence and IQ Magic in Education. A paper given at the Symposium, "Is the Construct of Intelligence a 20th Century Myth?" New York Psychological Association Meeting, 1979, 12p.

Horn, J.L. "Trends in the Measure of Intelligence." Intelligence, 1979, 3, 229-240.

Jensen, A.R. Bias in Mental Testing. New York: Free Press, 1980.

Jones, R., Wilderson, F. "Mainstreaming and the Minority Child: An Overview of Issues and Perspective." In R. Jones (ed.), Mainstreaming and the Minority Child. Reston, VA.: Council for Exceptional Children, 1976.

Kamin, L.J. The Science and Politics of IQ. New York: John Wiley and Sons, 1974.

Kaufman, A.S. Intelligence Testing with the WISC-R. New York: John Wiley and Sons, 1979.

Kaufman, A.S. "WISC-R Research: Implications for Interpretation." School Psychology Digest, 1979, 8, 5027.(b)

Kaufman, A.S., and Kaufman, N.L. Kaufman Assessment Battery for Children: Interpretative Manual. Circle Pines, Minn.: American Guidance Service, 1983.

Kennedy, W.A., Moon, H., Nelson, W., Lindner, R., and Turner, J. The Ceiling of the New Stanford Binet. Journal of Clinical Psychology, 1961, 17, 284-286.

Kirschenbaum, R.J. "Let's Cut Out the Cut-Off Score in the Identification of the Gifted." Roeper Review, 1983, May, 6-9.

Laosa, L.M. "Nonbiased Assessment of Children's Abilities." In Oakland (ed.), Psychological and Educational Assessment of Minority Children. New York: Brunner/Mazel, 1977.

Lennon, R.T. "Perspective on Intelligence Testing." NCME Measurement in Education, 1978, 9:2, 1-8.

Lennon, R. "The Anatomy of a Scholastic Aptitude Test." NCME Measurement in Education, 1980, 11, 1-8.

LeRose, B. "A Study of the Relationship between Psychometric and Piagetian Measures of Giftedness." In G. Lubin, M. Polusen, J. Magary, and M. Sota-McAlister (eds.), Piagetian Theory and Its Implications for the Helping Professions. Proceedings, Seventh Interdisciplinary Conference, Vol. 2. Los Angeles: University of Southern California, 1978.

Luria, A.R. (1966). Human Brain and Psychological Processes. New York: Harpert Row 1966.

Lyon, R. "Social and Legal Issues in Testing." In H.L. Swanson and B.L. Watson (eds.) Educational and Psychological Assessment of Exceptional Children. St. Louis: C.V. Mosby Co., 1983.

MacMillan, D. Mental Retardation in School and Society. Boston: Little Brown, 1977.

Mackay, A. "The Canadian Charter of Rights and Freedoms: Implications for Students." Paper delivered at National Educational Policy Conference. Simon Fraser University, Vancouver: June 1983.

Martinson, R.A. The Identification of the Gifted and Talented: An Instructional Syllabus for National Summer Leadership Training Institute on the Education of the Gifted and Talented. Ventura, California: Office of the Ventura County Superintendent of Schools, 1974.

Matheson, D.W. Confirmatory Factor Analysis of the WISC-R for Inuit Children. Unpublished Master's Thesis. University of Alberta, Fall, 1983.

Mercer, J. Labelling the Mentally Retarded. Berkeley, CA: University of California Press, 1973.

Mercer, J.R. "Latent Functions of Intelligence Testing in the Public Schools." In Latrar P. Millar (ed.), The Testing of Black Students. Englewood Cliffs, N.J.: Prentice-Hall, 1974, 77-94.

Mercer, J.R. "Psychological Assessment and the Rights of Children." In N. Hobbs (ed.), Issues in the Classification of Children (Vol. 1). San Francisco: Jossey-Bass, 1975.

Millar, M., and Walker, K. "The Myth of the L.D. WISC-R Profile." The Exceptional Child, 1981, 28:2, 83-88.

Mishra, S.P. "Cognitive Processes: Implications for Assessing Intelligence." Theory into Practice, 1980, 23:2, 145-150.

Mussen, P., Rosenzweig, M., Aronson, E., Elkind, D., Feshbach, S., Geuvitz, P., Glinckman, S., Murdock, B. Jr., and Wertheimer, M. Psychology: An Introduction. Lexington: D.C. Heath and Company, 1973.

Nader, R. "The Power and the Story of the Educational Testing Service." The Advocate, George Washington University, February 7th, 1979, 9.

Passout, A.H. "The Talent Among The Disadvantaged." Accent on Talent, 1966, 1, 3-7.

Patterson, J.L., Czajkowksi, T.J., Hubbard, E., Johnson, G., Slater, C., and Kaufman, D. "How to Avoid the Dangers of Testing." In P. Houts (ed.), The Myth of Measurability. New York: Hart Publishing Company, Inc., 1977.

Pegnato, C.W., and Birch, J.W. "Locating Gifted Children in Junior High Schools - A Comparison of Methods." Exceptional Children, 1959, 25, 300-304.

Perrone, V. "On Standardized Testing and Evaluation." In P. Houts (ed.), The Myth of Measurability. New York: Hart Publishing Company, Inc., 1977.

Peterson, D.R. The Clinical Study of Social Behavior. New York: Appleton-Century-Crofts, 1968.

Piaget, J. The Origins of Intelligence in Children. New York: Morton, 1952.

Piaget, J. The Theory of Stages in Cognitive Development. In D.R. Green (ed.), Measurement and Piaget. New York: McGraw-Hill, 1971.

Renzulli, J.S. "Talent Potential in Minority Group Students." Exceptional Children, 1973, March, 437-444.

Reschly, D.J. (1979). Nonbiased Assessment in G.D. Phye and D.J. Reschly (ed.). School Psychology Perspectives and Issues.

Reschly, D.J. "Psychological Testing in Educational Classification and Placement." American Psychologist, 1981, 36, 1094-1102.

Reynolds, C.R. "The Problem of Bias in Psychological Assessment." In C.R. Reynolds and T.B. Gutkin (eds.). The Handbook of School Psychology. New York: Wiley, 1982.

Rose, H., and Rose, S. "The IQ Myth." In D. Rubenstein, Education and Equality. London: Harper and Row, 1980.

Rosenham, D.L. "On Being Sane in Insane Places." Science, 1975, 179, 250-258.

Ross, S.L., DeYoung, H.G., and Cohen, J.C. "Confrontation: Special Education Placement and the Law." Exceptional Children, 1971, 38, 5-12.

Ryan, L.E., 1973. An Investigation of the Relationship Between the Scores Earned by Selected Negro and White Children on the Wechsler Intelligence Scale for Children and the Wide Merge Achievement Test (Doctoral Dissertation, Mississippi State University, 1973). Dissertation Abstracts International 1973, 34, 2398A-2399A University Microfilms No. 7325711.

Salvia, J., and Ysseldyke, J. Assessment in Special and Remedial Education. New Jersey: Houghton Mifflin, 1985.

Samuda, R. Psychological Testing of American Minorities: Issues and Consequences. New York: Dodd, Mead and Co., 1975.

Samuda, R.J. How are the Schools of Ontario Coping with a New Canadian Population? A Report of Recent Research Findings. Toronto: Ontario Ministry of Education, 1980.

Samuda, R.J., and Crawford, D.H. Testing, Assessment, Counselling and Placement of Ethnic Minority Students. Toronto: Ontario Ministry of Education, 1980.

Sapon-Shevin, M. "The Tug-of-War Nobody Wins: Allocation of Educational Resources for Handicapped, Gifted and 'Typical' Students." Curriculum Inquiry, 1984, 14:1, 57-79.

Sarasa, S. Psychological Problems in Mental Deficiency. (3rd Ed.) New York: Harper, 1959.

Sattler, J.M. Assessment of Children's Intelligence and Special Abilities. (2nd ed.) Boston: Allyn and Bacon Inc., 1982.

Schmidt, F.L., and Hunter, J.E. "Racial and Ethnic Bias in Psychological Tests." American Psychologist, 1974, 29, 1-8.

Smith, C.R., and Knoff, H.M. "School Psychology and Special Education Students' Placement Decisions: IQ Still Tips the Scale." The Journal of Special Education, 1981, 15, 55-64.

Staats, A.W. Social Behaviorism. Illinois: Dorsey Press, 1970.

Stephens, B., McLoughlin, J.A., Millar, C.K., and Glass, G.V. "Factorial Educational Measures and Piagetian Reasoning Assessments." Developmental Psychology, 1972, 6, 343-348.

Sternberg, R.J. "Lies We Live By: Misapplication of Tests in Identifying the Gifted." Gifted Child Quarterly, 1982, 26:4, 157-161.

Sternberg, B.J. "Human Intelligence." Phi Delta Kappan, 1984, June, 694-698.

The Task Force on Crosscultural Assessment. The Time for Action: Positions and Recommendations of the Task Force on Crosscultural Assessment. Northeast Regional Resource Centre, Hightstown, New Jersey, 1980, 33p.

Thurstone, L.L. "Primary Mental Abilities." Psychological Monographs, 1938, No. 1.

Turnbull, W.W. "Intelligence Testing in the Year 2000." Paper presented at the Annual Meeting of the American Educational Research Association 63rd, San Francisco, California, April, 8-12, 1979.

Vernon, P.E. The Structure of Human Abilities (2nd ed.). London: Methuen, 1961.

Vernon, P.E. "Intelligence and Attainment Tests." 1960.

Wallace, G., and Larsen, S.C. Educational Assessment of Learning Problems: Testing for Teaching. Boston: Allyn and Bacon, Inc., 1978.

Warren, S.A., and Brown, W.G. "Examiner Scoring Errors on Individual Intelligence Tests." Psychology in the Schools. 1973, 10, 118-122.

Wesman, A.G. "Intelligent Testing." American Psychologist, 1968, 23, 267-74.

Williams, R. "Abuses and Misuses in Testing Black Children." The Counselling Psychologist, 1971, 1, 62-73.

Wittshire, E.B., and Gray, J.E. "Draw-a-Man and Raven's Progressive Matrices (1938) Intelligence Test Performance of Reserve Indian Children." Canadian Journal of Behavioral Science, 1969, 1, 119-122.

Zigler, E. "Mental Retardation." Science, 1967, 157, 578-579.

Zimmerman, I.L., and Woo Sam, J.M. Clinical Interpretation of Wechsler Adult Intelligence Scale. New York: Gruen and Strutton, 1973.

Part 1 of the Constitution Act, 1982 (which is Schedule B of Canada Act, 1982 (U.K.)), 1982, c. 11.